

Project 1: MATLAB Automated Event Detection using Raw ECG Data

Jordan Hayes ... Grand Valley State University, School of Engineering

hayejord@mail.gvsu.edu, 989-667-3171

¹ Grand Valley State University

² Dr. Samhita Rhodes

Article Info

Word Count:

8 Figures and 2Tables

Received: October 28, 2021

EGR 434-01 Bioelectric Potentials

Abstract

Background and Aim: Modern cardiac arrhythmia detection suffers in its ability to reliably detect the presence of cardiac events and analyze their irregularities. Detection of the QRS complex is the first crucial step in all application of automatic Electrocardiogram (ECG) analysis. The purpose of this paper is to describe the development and performance of a QRS event detection algorithm based on the Pan-Tompkins model. Overall, simplicity and efficiency are required in developing QRS detection algorithms for processing long-term recordings and large databases.³

Materials and Methods: The project is comprised of two parts: data collection and data analysis. Data collection was performed using classical methods such as a transducer for sensing blood flow and electrodes for measuring ECG data. PowerLab hardware was configured with LabChart7 software. The data analysis was developed in MATLAB according to the Pan-Tompkins QRS detection algorithm. A bandpass filter preprocesses the signal to reduce interference, permitting the use of low amplitude thresholds in order to get high detection sensitivity.

Results: A QRS detection algorithm was developed and implemented in MATLAB. Detection is reliable in the combined dataset because of the use of filters, slopes, and amplitudes in the determination of a QRS complex. For the dataset available, the algorithm succeeded in detecting all 35 of the *R* peaks, and therefore the QRS events with a succession rate of 100%. The algorithm also correctly places all Q points. 34 of 35 widths were correctly estimated. The imperfection stems from an S point which resides in the post exercise portion of the dataset; their detection has a succession rate of 97.14%. Other parameters such as QRS width and average heart rate are also calculated with high accuracy.

Conclusions: The Pan-Tompkins method was extremely reliable when developed and tested against the same data set. The reliability of the algorithm in reference to the known dataset is very high, but the lack of rolling averages and stubborn noise may produce imperfections in detection when used on other subject data and applications. One notable aspect of this project was that the separate raw datasets were analyzed continuously, thereby emphasizing a rapid change in cardiac behavior when combining the two into one dataset, yet it still remained accurate. It can be predicted that this algorithm would be reliable for the majority of ECGs without extreme arrhythmia.

Introduction

The severe acute respiratory syndrome caused by coronavirus 2, also COVID-19, has brought many issues to light throughout the world's journey through the 20-21 pandemic. One of these examples is cardiovascular health. A number of infected people develop heart-related problems either spontaneously or as a complication of preexisting cardiac disease, which is the most prominent cause of death in the world. Development is then necessary in fields such as signal processing to help a new generation of portable Electrocardiogram (ECG) to be efficient and low power. The QRS complex is a combination of three of the graphic deflection seen on a typical ECG further dissected into segments illustrated in Figure 1.

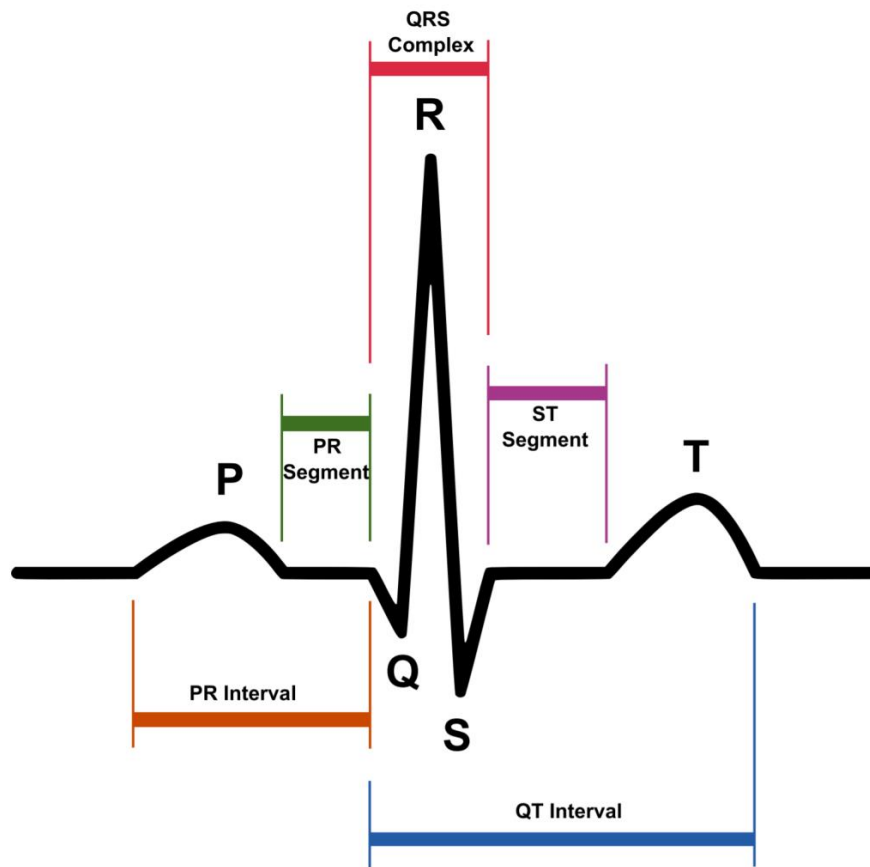


Figure 1: Main Events/Segments in ECG Signals

There are many different uses for algorithms that detect events such as these. It is especially necessary in analysis of cardiac health to correctly monitor and identify these events in cases of arrhythmia detection that can be as a result of a multitude of complex issues. The field of cardiovascular health has been heavily studied and intensely developed. Many algorithms have been created that use a variety of techniques such as applying a first derivative filter to the signal; this method is numerically efficient for the enhancement of certain phases of the QRS complex, but it is sensitive to noise and arrhythmia. Usually, denoising ECG signal requires a bandpass filter. More complex techniques can be applied to remove noise such as a Bayesian framework that filters better than a bandpass, but the method becomes numerically inefficient and would not perform at speeds required in real time. The purpose of this paper is to propose an automated algorithm with no user input to determine the position and duration of the QRS complexes using one lead of a standard 12-lead ECG monitoring system. The objectives will be completed by developing the algorithm in MATLAB to reliably and accurately detect the QRS complexes in the ECG data of subjects at rest and after one minute of exercise. The algorithm is developed and adapted according to the Pans-Tompkins method and interaction with raw ECG signals taken from a transducer and collected by PowerLab instruments. The algorithm will

accept a file containing raw ECG of any sample rate taken for 20 seconds. The intention is to test the algorithm against a data set consisting of 10 seconds of resting data and 10 seconds of post exercise data, display detection results, and calculate detection statistics. The development of signal processing algorithms is imperative to expanding our biomedical capabilities in the near future.

Methods

Selection and description of participants

The subjects under study in this test are EGR434-01, Bioelectric Potentials. students attending Grand Valley State University (GVSU). Subject range in age is negligible. 4 female and 2 male subjects participated in the collection of ECG data. Subjects do not have a history of cardiovascular or respiratory problems.

Data Collection

The raw ECG data was taken using a transducer to collect finger pulses and a Bio Amp with electrodes attached to the left wrist, right wrist, and right leg; an alternate electrode configuration may be used. Figures 2 and 3 show the hardware setup.

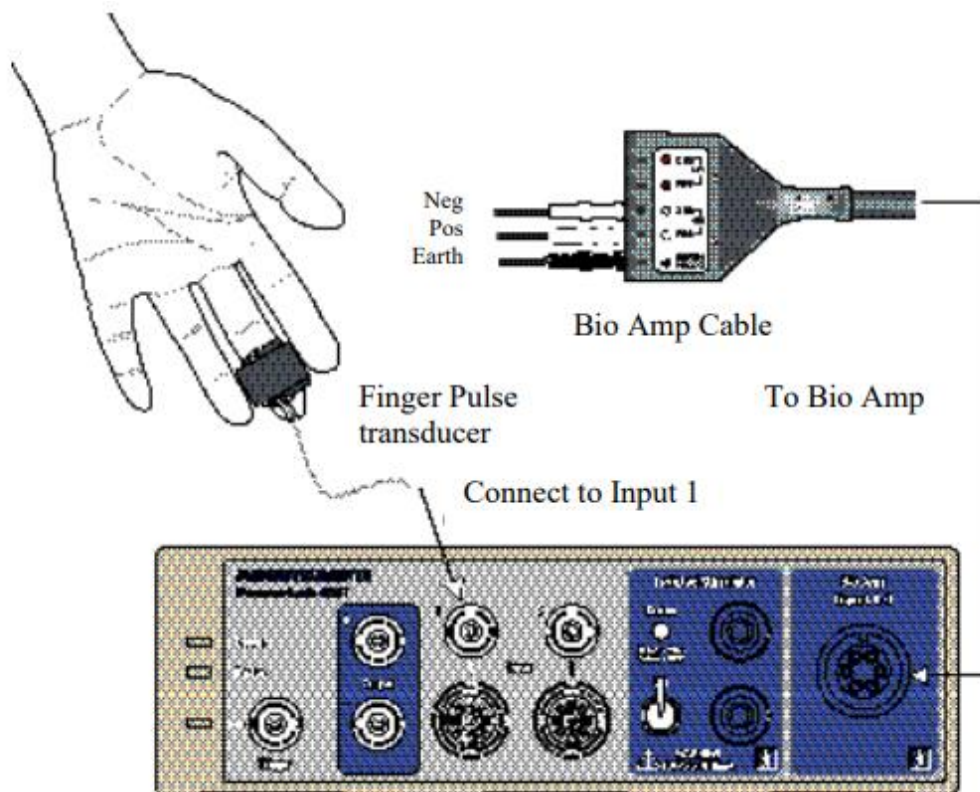


Figure 2: Experiment Hardware Connections to PowerLab

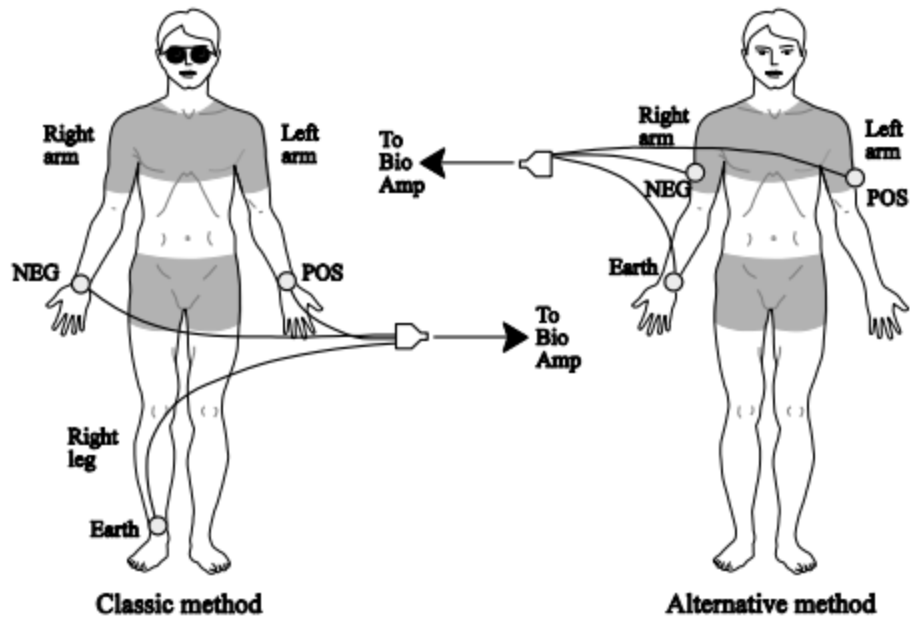


Figure 3: Configurations for Electrode Placement

LabChart7 software was used to communicate with the PowerLab equipment to log, display, and store raw bioelectric signals. Each signal was configured to a channel in LabChart. Channel settings were provided in a configuration file called “*ECG & Pulse Settings*”. Channel 2 is the raw signal from the finger pulse transducer and is an indication of the net rate of blood flow into the finger pulse. The time integral of Channel 2 is displayed on Channel 1 and illustrates the change in finger pulse volume over time. Channel 3 illustrates the ECG signal. The volunteer was asked to remain as relaxed and still as possible to minimize any affects of movement on the signal. The completed collection of raw data is shown in Figure 4.

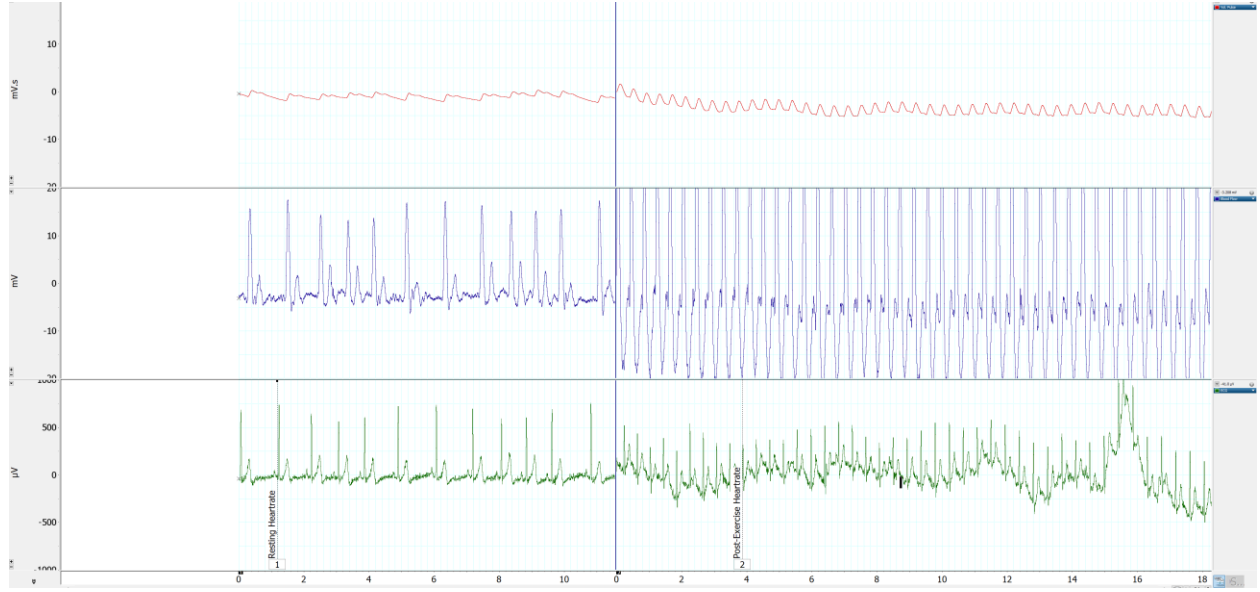


Figure 4: Raw Bioelectric Data vs Time Collected and Displayed By LabChart Software

The data was then exported as a Unicode Text File and copied into Microsoft Excel for removal of unused logged data and extension of time vector to include both sets as continuous time. The result is a 20 second vector containing 10 seconds of resting and 10 seconds of post exercise data.

Algorithm Design and Data Analysis

The Pan-Tompkins QRS Detection Algorithm was chosen for the development of processing the data above because of its numerical efficiency with the processing power available personal familiarity with signal processing methods used. Overall, simplicity and efficiency are required in developing QRS detection algorithms for processing long-term recordings and large databases.³ The signal is passed through a sequence of processing steps including three digital filters. The analytical methods are described and derived below.

The first objective was to remove noise using a bandpass filter composed of cascaded low-pass (LP) and high-pass (HP) filters. The transfer function of the second-order LP filter is shown in Eq. (1). Each filter will have a processing delay. The LP filter has a filter processing delay of 6 samples. Each filter was applied using convolution; in MATLAB this is the *conv()* function.

$$H(z) = \frac{(1-z^{-6})^2}{(1-z^{-1})^2} \quad (1)$$

The design of the HP filter is aimed to subtract the output of a first order LP filter from an all-pass filter. The transfer function for the HP filter is shown in Eq. (2). The HP filter has a signal processing delay of 16 samples.

$$H(z) = \frac{(-1+32z^{-16}+z^{-32})^2}{(1-z^{-1})} \quad (2)$$

The resulting signal is passed through a derivative filter to provide the QRS complex slope information. This will help determine slopes that are prominent enough to be analyzed as part of the QRS complex. A five-point derivative filter was used with a transfer function described by Eq. (3). Its delay is two samples.

$$H(z) = \left(\frac{1}{8}T\right)(-z^{-2} - 2z^{-1} + 2z^1 + z^3) \quad (3)$$

The resulting signal is then squared point by point.

$$y(nT) = [x(nT)]^2 \quad (4)$$

After squaring, the signal is passed through a moving window integrator to calculate adaptive thresholds that discriminate the location of the QRS complexes. This is done because of irregularities in QRS complex shape and magnitude the heart produces under different scenarios, such as the resting and post exercise data were taken. This gives additional waveform feature information.

$$y(nT) = \frac{1}{N} [x(nT - (N - 1)T) + x(nT - (N - 2)T) + \dots + x(nT)] \quad (5)$$

The number of samples N in the moving window is important. For a sample rate of 200 samples/s, the window is 30 samples wide (150 ms). If the window is too wide, the integration waveform will not be able to distinguish the QRS from the T complexes. If it is too narrow, some QRS complexes will produce more than one peak inside the integration window. These can cause difficulty in subsequent QRS detection processes.⁴

The QRS complex now corresponds to rising edges of the integration waveform, and the time duration of the rising edge is equal to the width of the QRS complex. The algorithm takes advantage of the newly emphasized peaks, shape, and normalization. The max peak of the processed signal is found using MATLAB function *max()*. The threshold for determining the peaks is the mean of the processed signal; this mean represents the percentage of the max peak required for a voltage to be considered as an R wave and attempts to consider irregular behaviors. The MATLAB function *findpeaks()* was used to apply the threshold as well as a minimum peak separation to the processed signal to avoid marking potentially high T waves or noise. A binary vector was created to represent the indices of the measured voltages above threshold. MATLAB capabilities were then used to determine at which indices the binary vector changed from 0 to 1 and 1 to 0 representing rising and falling edges, respectively, which correspond to a window of samples in the time domain. The index of the desired points of the QRS complex were then determined. The maximum ECG value in each window was found using *max()*. The locations are marked as the R peaks. The Q point is then the minimum ECG value between the rising edge and the R peak. The S point is the minimum ECG value between the R

peak and the falling edge. The values and locations of the QRS complex characteristics were recorded and plotted over the normalized time domain signal.

The widths of the QRS complexes were estimated using the difference in number of samples from the S and Q points corresponding to the detected R peak, multiplied by the sampling period.

$$Width_{QRS} = (xVal(S) - xVal(Q)) * T_s \quad (6)$$

An average heartrate was calculated for the first ten seconds, the next ten seconds, and for the total ECG vector. The average was found in MATLAB by finding the number of peaks detected in first 10 and last seconds in the ECG vector, and dividing the number by time in minutes. The number of peaks in the first and last half of the ECG signal were added to calculate the overall average.

$$Avg_{bpm} = \frac{peaks}{time} \quad (7)$$

It should be noted that the signal is normalized after being processed by each filter for convenience.

$$y(t) = \frac{x(t)}{\max(abs(x(t)))} \quad (8)$$

Performance Analysis

The performance of QRS detection algorithms are typically assessed using two statistical measures: sensitivity (SE) and positive predictivity (+P) ³.

$$SE = \frac{TP}{TP+FN} \quad (9)$$

$$+P = \frac{TP}{TP+FP} \quad (10)$$

where TP is the number of true positive QRS detected, FN is false negatives which are QRS complexes which have not been detected, and FP is the number of false positives.

Results

The results of the loading of the data into MATLAB is shown in Figure 5.

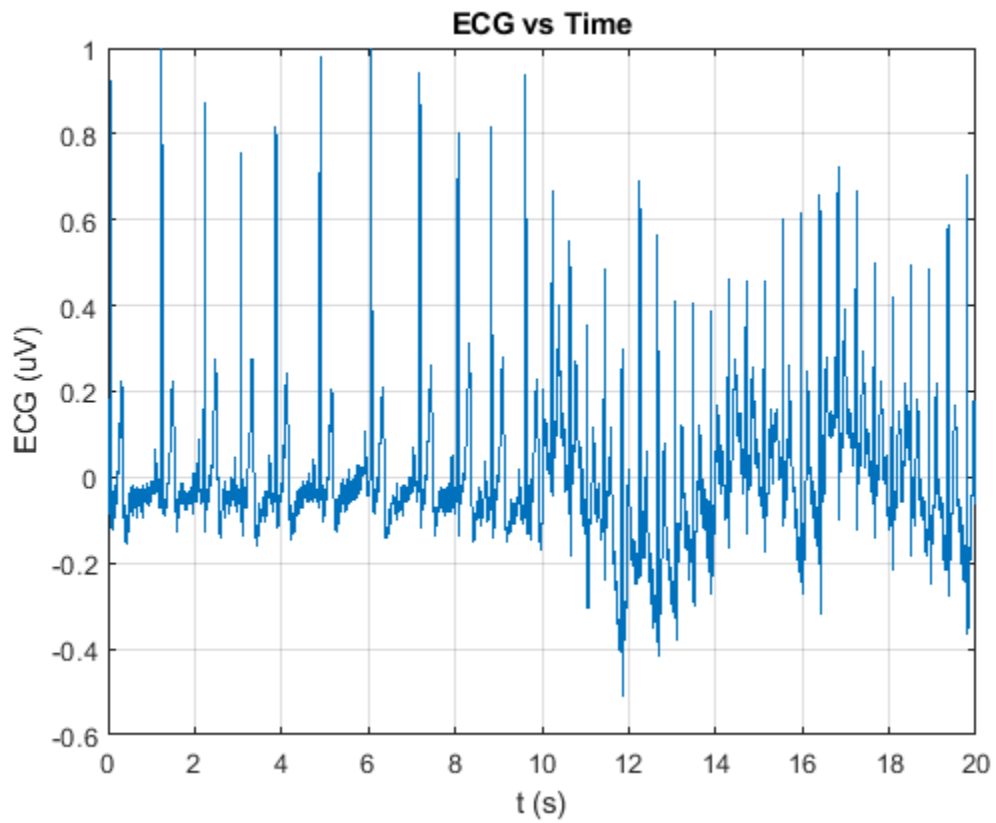


Figure 5: Raw Combined Resting and Post Exercise ECG vs. Time

The estimated indices and magnitudes of the QRS points were plotted over the same data for visualization.

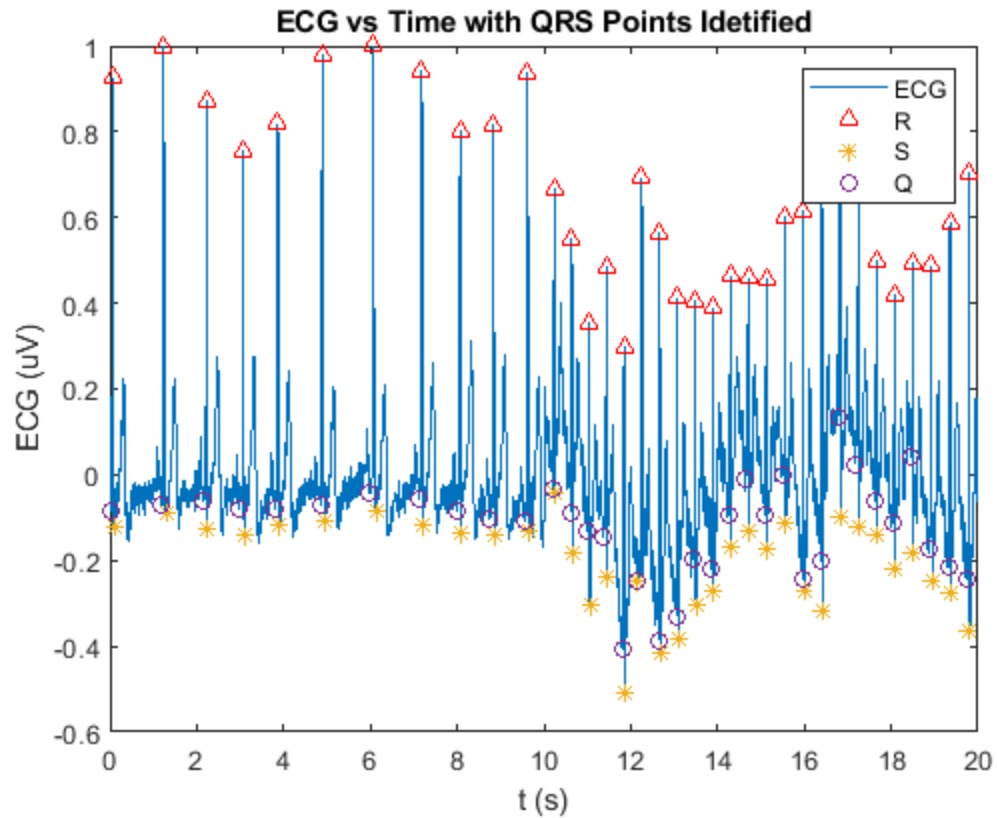


Figure 6: QRS Detection Over Raw ECG Data

Studying the plot shows that the *R* indicators are correctly centered at the peaks. The *Q* indicators come before the *S* except at 12 seconds, where the complex is extremely thin. The widths are further examined below.

Table 1: Widths of QRS complexes

QRS Event Number	Width (s)
1	0.08
2	0.14
3	0.115
4	0.115
5	0.095
6	0.08
7	0.165
8	0.085
9	0.095
10	0.105
11	0.085
12	0.055
13	0.05
14	0.08
15	0.095
16	0.05
17	0
18	0.055
19	0.05
20	0.085
21	0.075
22	0.065
23	0.08
24	0.06
25	0.045
26	0.045
27	0.05
28	0.04
29	0.08
30	0.065
31	0.05
32	0.065
33	0.09
34	0.045
35	0.09

The table shows all QRS complex widths calculated by the algorithm; all are positive, however, peak 17 has a width of 0. A width was calculated for all 35 peaks found. Widths 1-11 are generally larger compared to the rest.

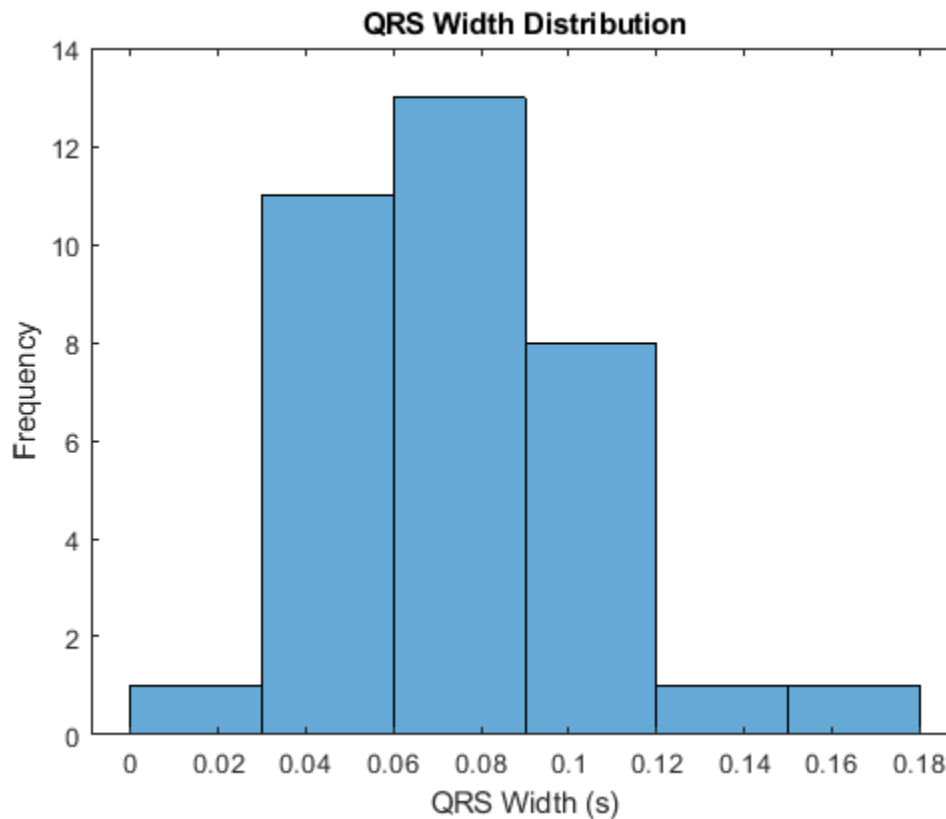


Figure 7: Histogram for QRS Complex Widths

The average heartrates for resting, post exercise, and combined datasets was calculated and the MATLAB output is shown below. There are more QRS widths to the left of the median.

The average resting heartrate is 66.00
 The average post exercise heartrate is 144.00,
 The overall average heartrate is 105.00

Figure 8: MATLAB Heartrate Estimations

From the MATLAB output, heartrate more than doubles in frequency when data transitions from resting to post exercise data. The average overall heartrate is always centered in between them.

The Pan-Tompkins method was extremely reliable when developed and tested against the same data set. Performance analysis was performed on the results of the detection. The total complex points to find was 105, 35 points for *Q*, *R*, and *S*. The algorithm identified 100% of QRS complexes through *R* peaks and a 99.67% detection rate for all points, with 100% succession for *Q* and *R* points, and 97.14% for *S*.

Table 2: Performance Analysis On QRS Complex Components

QRS Component	SE (%)	+P (%)
Q	1	1
R	1	1
S	97.14	97.14
Total	99.05	99.05

The performance parameters happen to be the same for each QRS point.

Discussion

The development of signal processing algorithms is imperative to the progress of detecting, analyzing, and diagnosing cardiac behavior. Detection is reliable in the datasets because of the use of signal processing techniques such as filters, slopes, and amplitudes in the determination of a true QRS complex. For the dataset available, the algorithm succeeded in detecting all 35 of the *R* peaks, and therefore the QRS events with a succession rate of 100%. The program also correctly identified 100% of *Q* points. In the case of *S* identified 34 of the 35 QRS events, a succession rate of 97.14%. Peak 17 shows an imperfection in the performance of the algorithm, as the same index was identified as both a *Q* and *S* point, resulting in a width of zero; the misidentified *S* segment resided in the exercise portion of the data. The reason performance parameters were the same is because the mistake in *S* identification was a false positive at the *Q* point it was identified as, and a false negative at the theoretical index. Other parameters such as QRS width and average heart rate are also calculated with high accuracy that described the reaction of heartrate and QRS complex to exercise. The sensitivity reports the percentage of true beats that were correctly detected by the algorithm; the positive predictivity reports the percentage of beat detections that were true beats.⁵

There are more QRS widths to the left of the median. Measured QRS widths 1-11 displayed in Table 1 are generally larger compared to the rest because resting data has 11 peaks, shown by Figure 7. Resting widths consistently fall between .09 and .12. Then heartrate and QRS shape immediately change as it moves into the post exercise data, where post exercise QRS widths have more frequency and higher variability. After exercise, heartrate increases and therefore so does the number of peaks and widths. The *QS* width shortens and becomes more variable as *R* peaks fluctuate. From heartrate calculations, it can be seen that an increase in heartrate corresponds to a decrease in QRS width and an increase in frequency. From the MATLAB output, it can be seen that the heartrate more than doubles in frequency when data transitions

from resting to post exercise data. The average overall heartrate is always centered in between them because of software operations.

The reliability of the algorithm in reference to the known dataset is very high, but the lack of rolling averages and stubborn noise may produce imperfections in detection when used on other subject data and applications with more drastic differences from resting heartrate behavior and more noise that the signal floats over. Adding a variable threshold calculation would improve the accuracy of complex identification with more variability in physiology. One notable aspect in the development of this project was that the raw data was analyzed continuously even though data was not taken continuously, thereby emphasizing a rapid change in cardiac behavior. The combination of the two data sets do not represent linear time, as the subject has exercised unmonitored by data collection; instead, this was done in order to improve the robustness to noise and irregular heart beats. It can be predicted that this algorithm would be reliable for the majority of ECGs without extreme arrhythmia.

Tables

Table 1: Widths of QRS complexes

QRS Event Number	Width (s)
1	0.08
2	0.14
3	0.115
4	0.115
5	0.095
6	0.08
7	0.165
8	0.085
9	0.095
10	0.105
11	0.085
12	0.055
13	0.05
14	0.08
15	0.095
16	0.05
17	0
18	0.055
19	0.05
20	0.085
21	0.075
22	0.065
23	0.08
24	0.06
25	0.045
26	0.045
27	0.05
28	0.04
29	0.08
30	0.065
31	0.05
32	0.065
33	0.09
34	0.045
35	0.09

Table 2: Performance Analysis On QRS Complex Components

QRS Component	SE (%)	+P (%)
Q	1	1
R	1	1
S	97.14	97.14
Total	99.67	99.

Figures

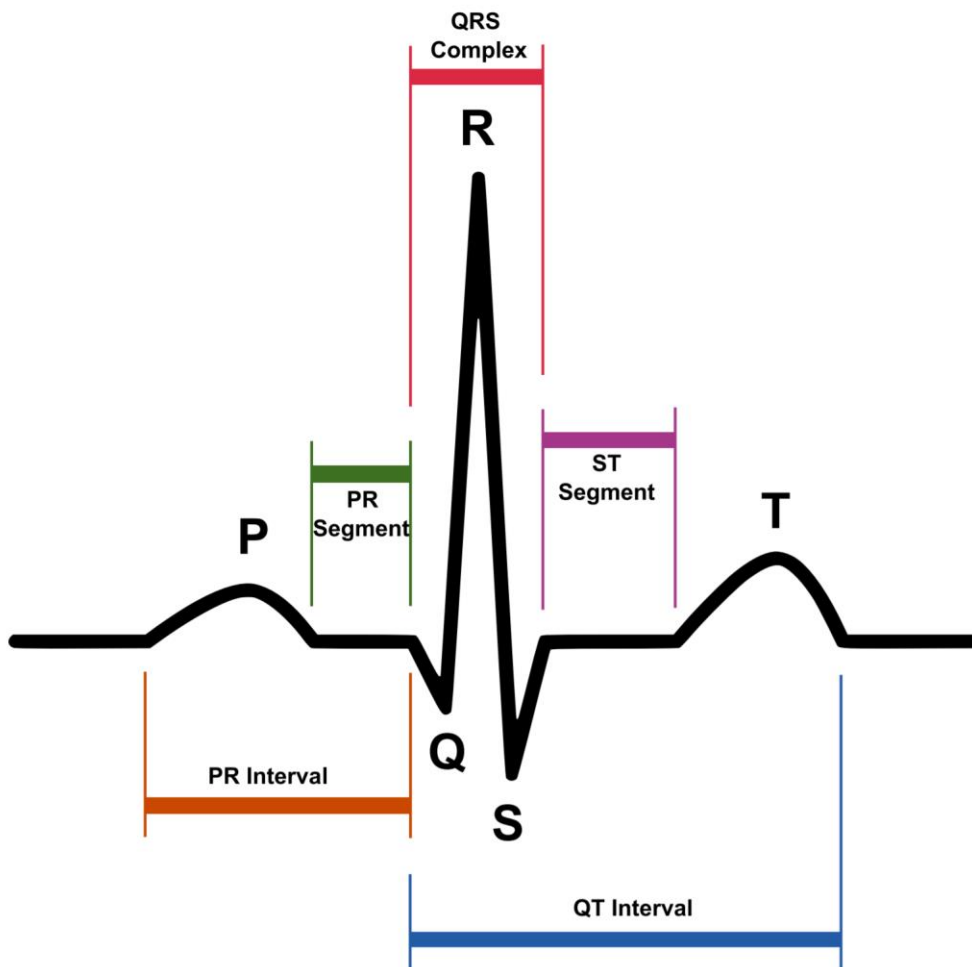


Figure 1: Main Events/Segments in ECG Signals

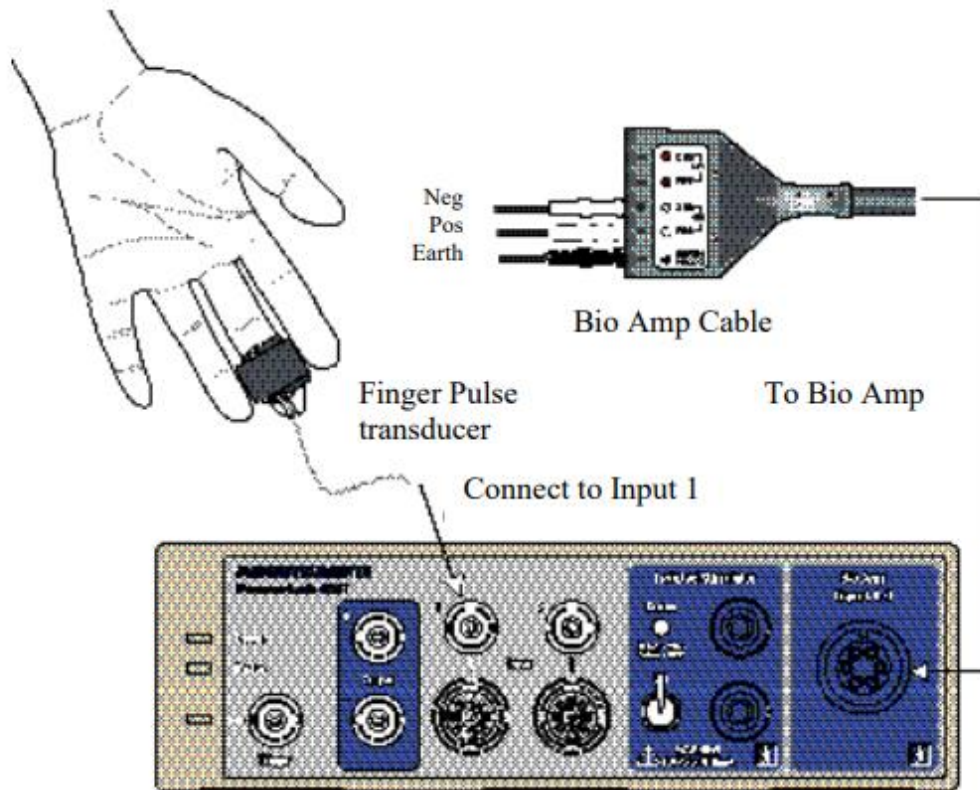


Figure 2: Experiment Hardware Connections to PowerLab

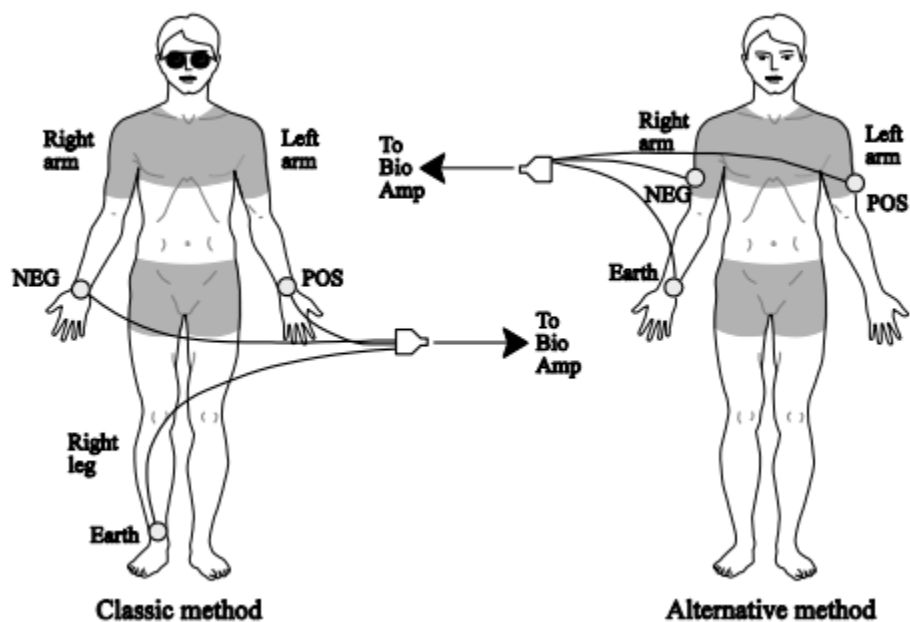


Figure 3: Configurations for Electrode Placement

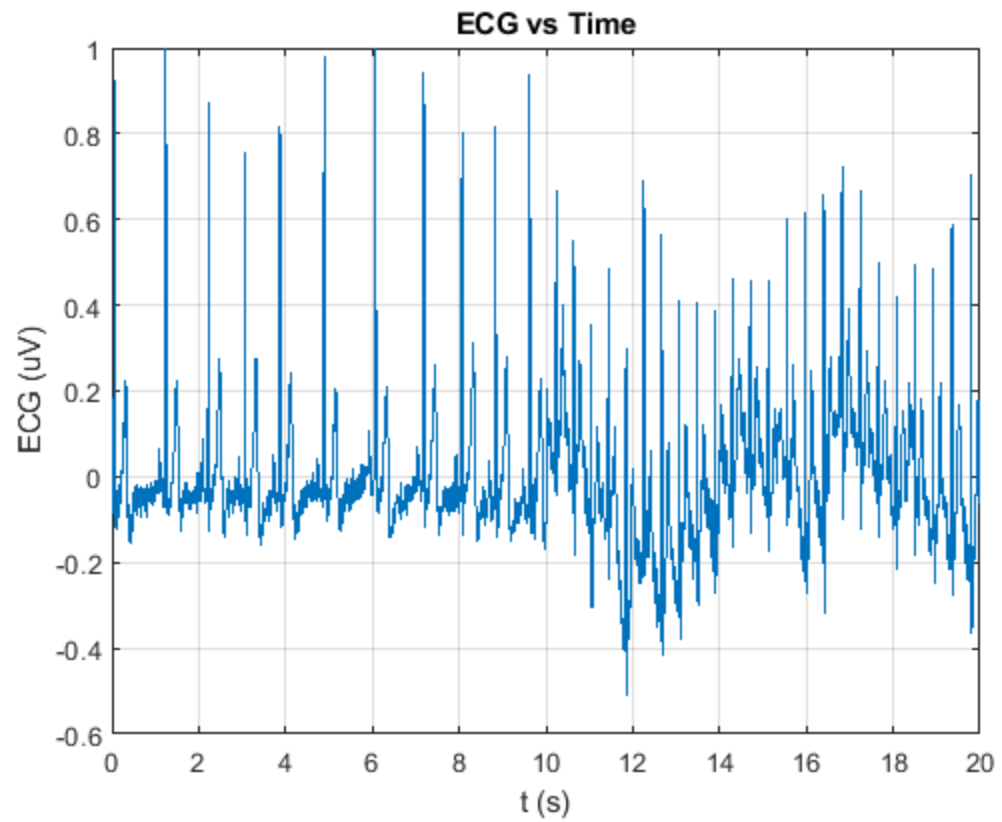


Figure 5: Raw Combined Resting and Post Exercise ECG vs. Time

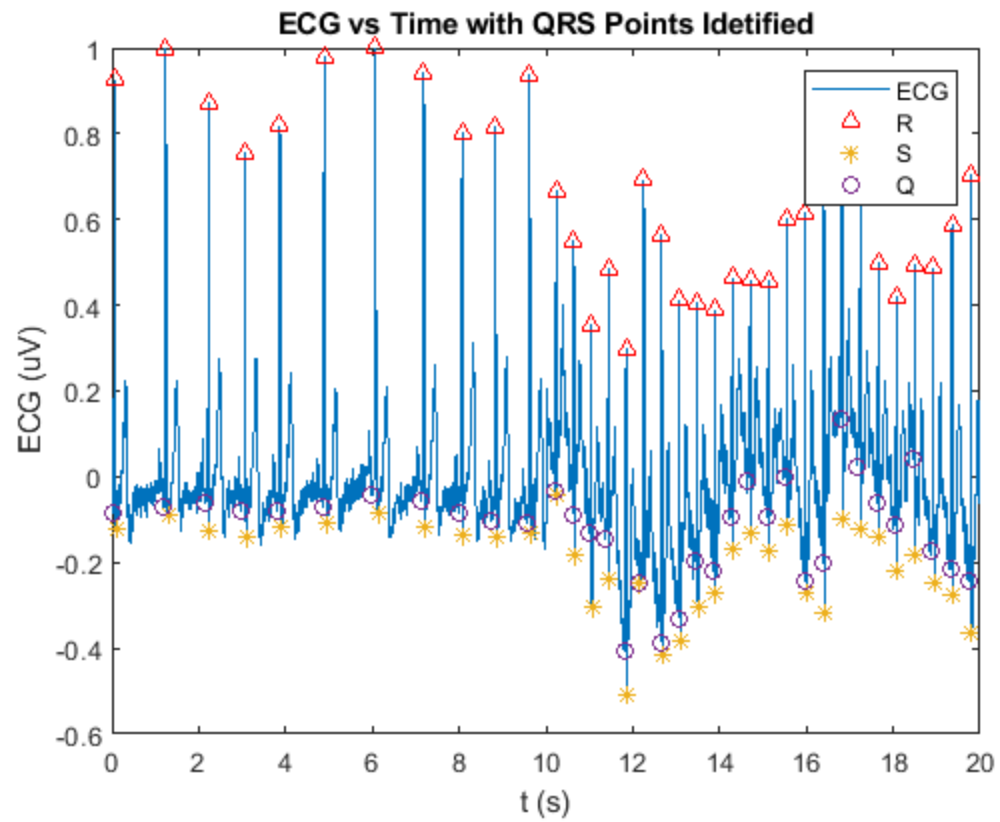


Figure 6: QRS Detection Over Raw ECG Data

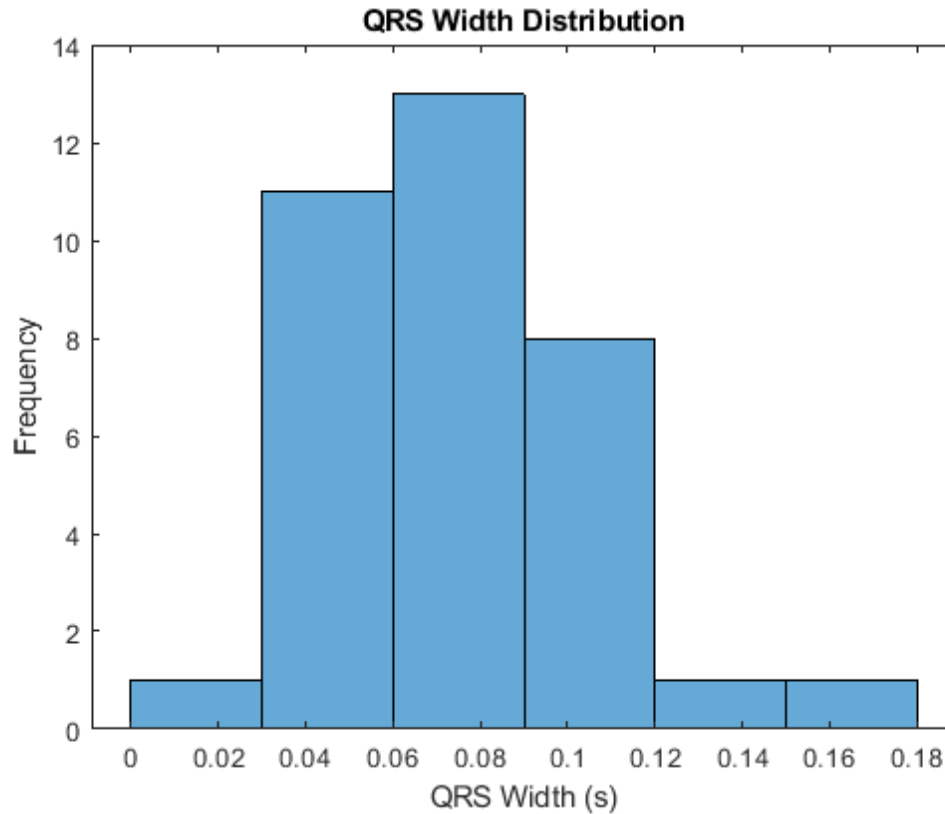


Figure 7: Histogram for QRS Complex Widths

The average resting heartrate is 66.00
The average post exercise heartrate is 144.00,
The overall average heartrate is 105.00

Figure 8: MATLAB Heartrate Estimations

Acknowledgements

Authors would like to acknowledge the GVSU School of Engineering for providing the equipment and software needed for the completion of the project, as well as Dr. Samhita Rhodes, Instructor for EGR 434-01 Bioelectric Potentials, for critical feedback and technical expertise.

References

- Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Transactions on Biomedical Engineering. 1985;BME-32(3):230–6.
- Elgendi M, Eskofier B, Dokos S, Abbott D. Revisiting QRS detection methodologies for Portable, wearable, battery-operated, and wireless ECG systems. PLoS ONE. 2014;9(1).
- Uysal F. QRS Detection Using Pan-Tompkins algorithm [Internet]. OpenStax CNX. [cited 2021Oct25]. Available from: https://cnx.org/contents/YR1BU9_@1/QRS-Detection-Using-Pan-Tompkins-algorithm

Endnotes

- ¹ Equipment and software provided by: Grand Valley State University
- ² Instructor for EGR 434-01 Bioelectric Potentials
- ³ Elgendi, Eskofier, et al. (p. 15)
- ⁴ Pan, Tompkins (p. 233)
- ⁵ Elgendi, Eskofier. et al. (p. 12)